

# 202

## Reliability and Validity Levels of Measurement



# True score vs. measurement error

Causes of all the problems in the world



- True score: objective truth
- Observed score: what you observe
- Measurement error: difference between them

# **The Goal of Measurement: Validity and Reliability**



# Reliability

- The degree to which our measurements are consistent and contain a minimum of measurement error
  - If you weigh yourself multiple times during the day, do you get the same result?

# Types of Reliability

## INTERNAL

(extent to which a measure is consistent within itself.)

---

### **split-half method:**

measures the extent to which all parts of the test contribute equally to what is being measured.

## EXTERNAL

(the extent to which a measure varies from one use to another.)

---

**test re-test:** measures the stability of a test over time.

---

**Inter-rater:** to the degree to which different raters give consistent estimates of the same behavior

# Internal Reliability

People's responses should be consistent if the measure reflects the same underlying construct

Example: Rosenberg self-esteem scale

- People who agree that they are a person of worth should report that they have good qualities
- Vice versa

no, you're not perfect.



and that's exactly how it should be.



# Types of Reliability

- Test-retest
- Inter-rater
- Internal consistency
  - Chronbach's alpha
  - Split-half reliability
- Assess reliability with correlation ( $r$ )



# Split Half (Internal)

## Test-Retest Reliability (External)

Half 1

Half 2

Time 1

Time 2

50

60

60

75

75

80

40

50

30

40

45

55

77

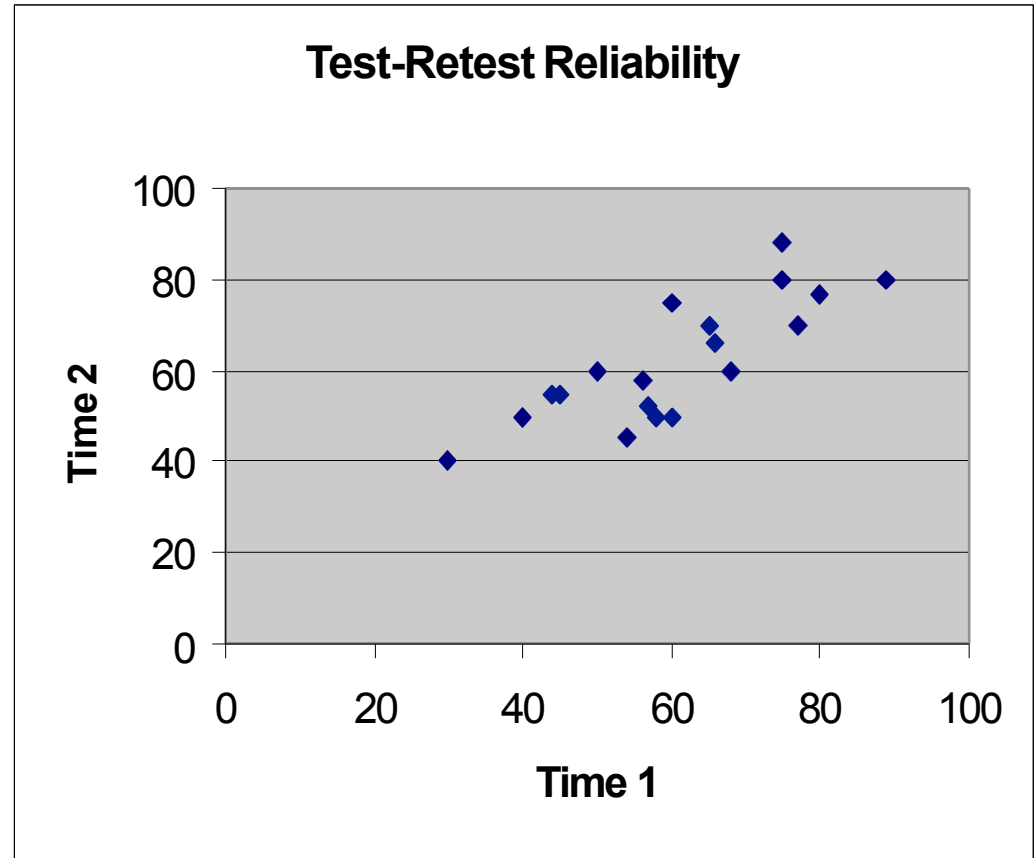
70

89

80

58

50

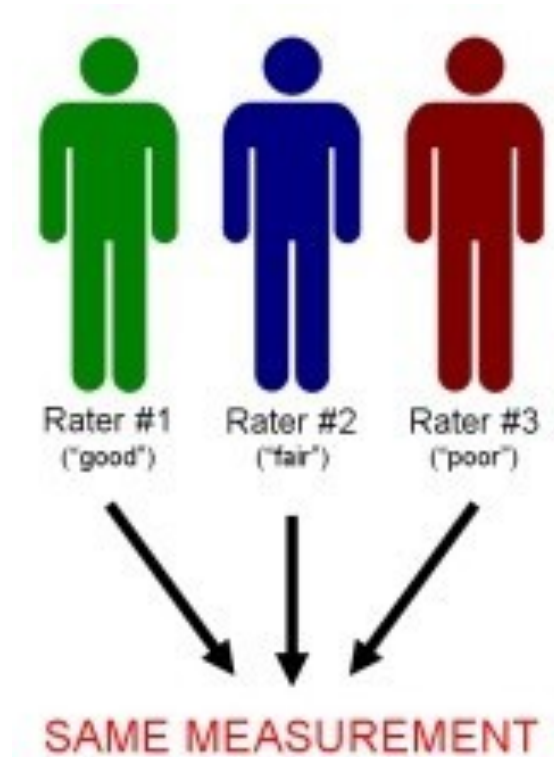




# Test-Retest, Split-Half Reliability

- Correlation Coefficient
  - Direction + or -
  - Degree 0 to 1  
-1 to 0
- Pearson Product Moment
  - $r = .82, n = 40, p < .05$

# Interrater reliability



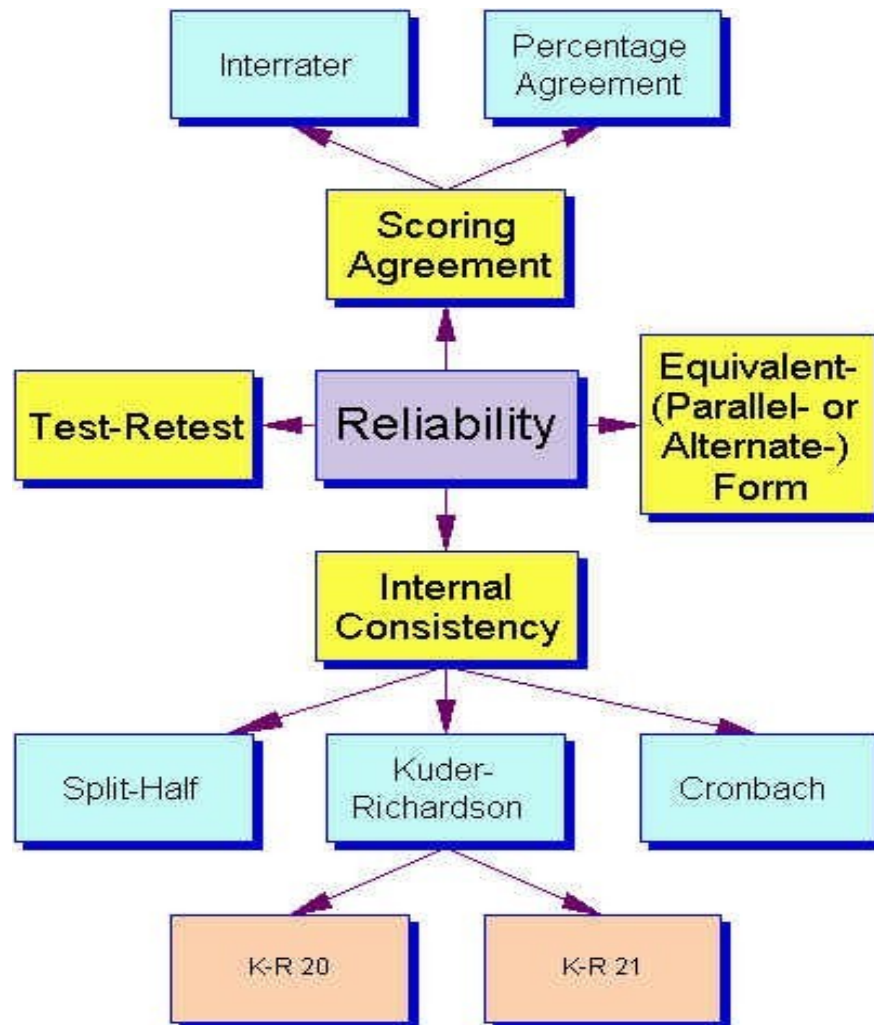
# Internal Consistency Test with Chronbach's alpha



## Internal consistency reliability (cont.)

- A questionnaire is administered to 1 group of subjects on 1 occasion
- The results are examined to see how well questions correlate
- If reliable, each question contributes in a similar way to the questionnaire's overall score

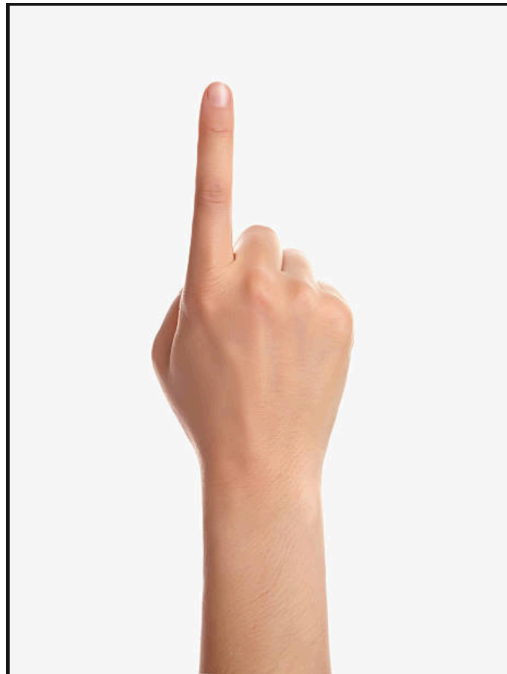
# Reliability



# Is reliability enough?

What if someone believes the length of your index finger reflects their self esteem?

- High internal and external reliability
- Is this really measuring self esteem?



# Validity

- The extent to which a procedure/instrument measures what it is intended to measure.





# Validity

## ♣ Construct validity

- the adequacy of the operational definition(s)

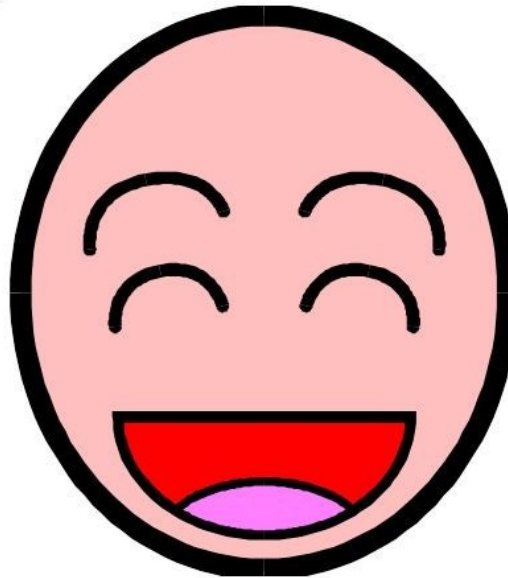
## ♣ Internal validity

- the ability to draw conclusions about causal relationships

## ♣ External validity

- the extent to which the results can be generalized to other populations and settings

# Face Validity



Looks Good To Me

- It looks OK
- Looks to measure what it is supposed to measure
- Look at items for appropriateness
  - Client
  - Sample respondents
- Least scientific validity measure

# Content Validity

Does measure assess appropriate content?

## Content validity

Spider phobia

### Aspects of the construct

Strength of fear reaction

Persistence of reaction

Invariability of reaction

Recognition that reaction is unreasonable

Avoidance of spiders

...

### Aspects assessed



# Criterion Validity (Crit)

Do the scores correlate with other variables you would expect them to correlate to?

Example: test anxiety

- Should negatively correlate with test performance
- If not, it may be due to your questionnaire, rather than a novel finding



# Predictive Validity (Crit)

Does measure predict future behavior?

## Predictive Validity

Does the test predict  
something  
useful like?



on-the-job performance?

# Convergent Validity

Are scores related to similar measures?

Need for Cognition Scale – used to assess how much people value and engage in thinking

- Highly correlated with standardized tests
- Negatively correlated with dogmatism



# Discriminant Validity

Scores not related to different measures

- Bipolar disorder, schizophrenia and normal controls
- Self-esteem (stable) vs. mood (unstable)

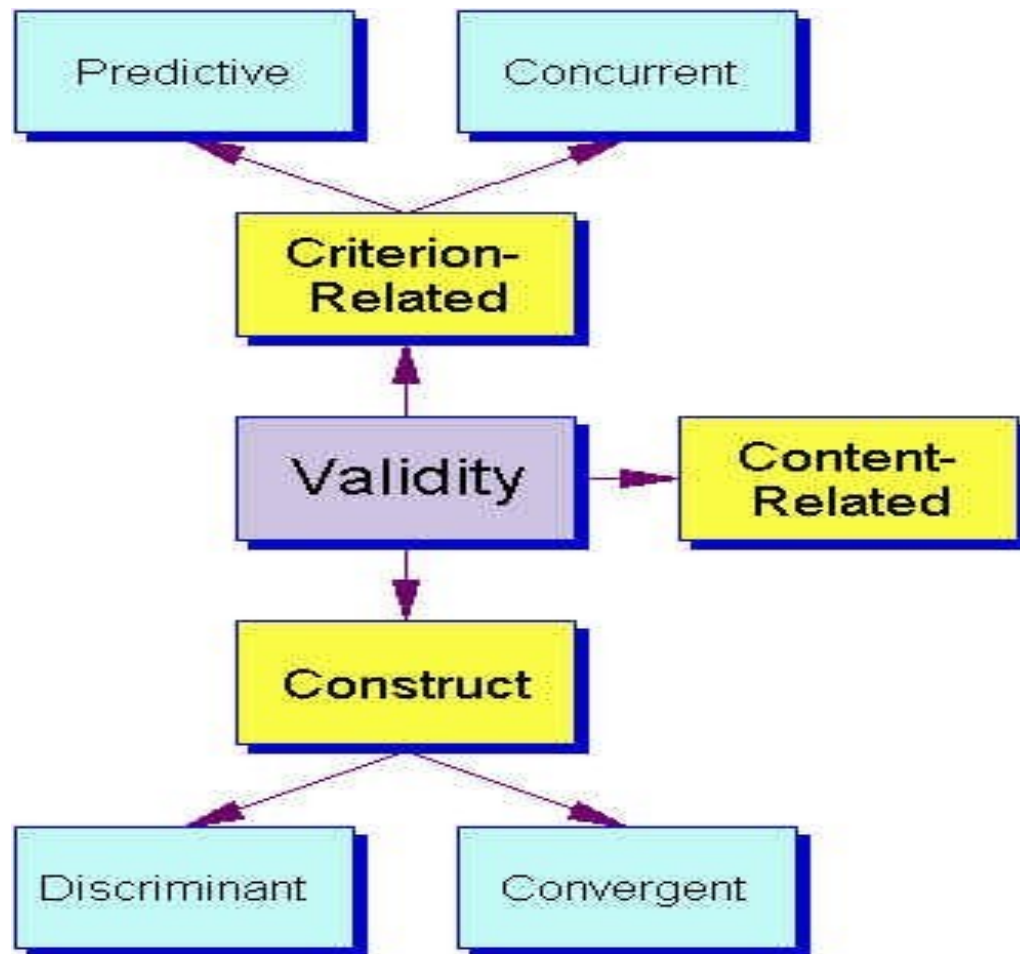


# Convergent vs. Discriminant

## **Convergent and Discriminant Validity**

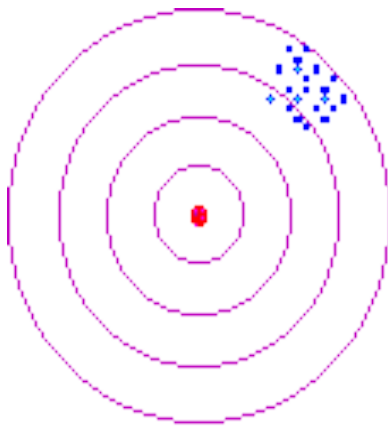
- To have construct validity, a measure should both:
- Correlate with other measures that it should be related to (**convergent validity**)
- And, not correlate with measures that it should not correlate with (**discriminant validity**)

# Validity

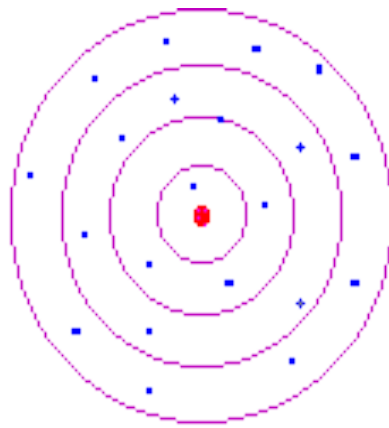


# Reliability and Validity

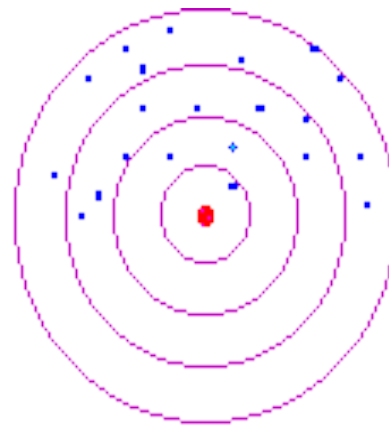
Think of the center of the target as the concept that you are trying to measure.



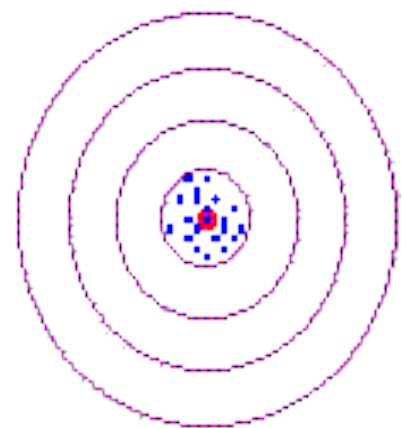
**Reliable  
Not Valid**



**Valid  
Not Reliable**



**Neither Reliable  
Nor Valid**



**Both Reliable  
And Valid**

# Think, pair, share

Think back to the last college exam you took and think of the exam as a psychological measure.

- What construct do you think it was intended to measure? Comment on its face and content validity.
- What data could you collect to assess its reliability and criterion validity?

# Questions to ask in selecting measures for a study

- Use standardized or new measures?
- What about reactivity?
- Do the scores reflect error?
- Do scores reflect the hypothetical construct?
- Will the results generalize beyond our study?

# Reactivity

- Does awareness of being measured change a participant's behavior?



# Levels of Measurement

- Nominal
- Ordinal
- Interval
- Ratio



# Levels

## Non-parametric

- Nominal
- Categorical

## Parametric

- Interval/Ratio
- Continuous
- Scale

# Personality Tests

## The Five Factor Model of Personality



		Low Scorers	High Scorers
1	Extraversion	Lonely Quiet Passive Reserved	Lonely Talkative Active Affectionate
2	Agreeableness	Suspicious Critical Ruthless Irritable	Trusting Lenient Soft-hearted Good-natured
3	Conscientiousness	Negligent Lazy Disorganized Late	Conscientious Hard-working Well-organized Punctual
4	Neuroticism	Calm Even-tempered Comfortable Unemotional	Worried Temperamental Self-conscious Emotional
5	Openness to experience	Down-to-earth Uncreative Conventional Uncurious	Imaginative Creative Original Curious